

## Corrigé : TP non supervisé, analyse exploratoire

### Solution: Exercice 1

1. un axe nord/sud et un axe est/ouest
2. La température annuelle est la plus liée à l'axe 1 (nord/sud), et tous les mois sont positivement corrélés (effet taille), mais surtout ceux des saisons intermédiaires. C'est l'amplitude qui est la plus corrélée à l'axe 2 (climat continental / océanique).
3. 98.3% de la variance est expliquée par les 2 premiers axes. Il est inutile de considérer les axes suivants.
4. Individus actifs représentatifs : axe 1 : Athens, Elsinke ; Axe 2 : Dublin / Kiev ou Budapest.
5. Individus supplémentaires représentatifs : Axe 1 : Seville / St Petersburg, Axe 2 : St Petersburg/ Edinburg
6. Mois contribuant le plus à l'inertie :  
axe 1 : Avril ,septembre, octobre ; axe 2 : Juin, Janvier.
7. les axes auxquelles se rattachent les variables quantitatives supplémentaires sont respectivement : annual : 1, amplitude :2 , latitude :1, longitude :2.
8. Variable qualitative :  
East : 2, positif. West : 2, négatif. North : 1, négatif. South : 1, positif.
- 9.

```
res$var$coord[1,c(1,2)]
res$quanti.sup$coord[2,c(1,2)]
```

Janvier est positif sur la composante 1 et négatif sur la composante 2. C'est l'inverse pour Amplitude.

→ le produit scalaire des 2 variables est négatif donc la corrélation est négative (si les deux premiers axes suffisent à représenter correctement les données).

confirmation :

```
cor(temperature[1:23,1:16]) # corrélation de -0.76
```

Les pays avec hiver comparativement doux ont moins d'amplitude (océaniques)

Brève conclusion : les oppositions des climats en Europe se résument entre chaud/froid (= sud/nord) et océanique/continental (ouest/est)

### Solution: Exercice 2 1.

```
min(which(resJO$eig[,3]>=50))
```

Il faudrait 6 axes pour expliquer 50% de la variance.

2.

```
t(rowW) %*%resJO$row$coord
```

3. Inertie sur le premier axe :

```
sum(t(rowW) * resJO$row$coord[,1]^2 )
resJO$eig[1,1]
```

4. Les coordonnées des profils sur les axes sont les composantes principales, qui sont orthogonales. la corrélation entre les vecteurs de coordonnées est donc nulle.

```
sum(t(rowW) * resJO$row$coord[,1] * resJO$row$coord[,2])
```

5. Contributions des lignes aux axes :

```
rowW * resJO$row$coord[,1]^2 / resJO$eig[1,1]
sum(rowW * resJO$row$coord[,2]^2 / resJO$eig[2,1])
### retrouver le resultat avec resJO$row$contrib
rowW * resJO$row$coord[,1]^2 / resJO$eig[1,1] -
  resJO$row$contrib[,1]/100
```

6. interprétation

- (a) axe 1 : endurance contre le reste (avec gradient pour l'endurance)  
axe 2 : lancers et marche contre sprint (avec gradient pour le sprint).

- (b)

```
ctrs <- colW * resJO$col$coord[,1]^2/resJO$eig[1,1]
ctrs - resJO$col$contrib[,1]/100
```

- (c) les pays ayant remporté le plus de médailles en endurance / autres :

ken	eth	mar	usa	gbr
-----	-----	-----	-----	-----

- (d) est-ce que usa est du côté de endurance ou de 'autres' ?

```
which(colnames(JO)=='usa')
#' 57
resJO$col$coord[57,]
# '--> autres
#' idem avec gbr
which(colnames(JO)=='gbr')
resJO$col$coord[23,]
#' --> autres
```

- (e) Conclusion axe 1 : endurance vs reste, axe 2 : sprint vs reste, à la fois pour les disciplines et les pays.

### Solution: Exercice 3 1. Représenter les données pour chacune des variables.

```
plot(banque$age)
plot(banque$sexe)
plot(banque$interdit)
plot(banque$credhab)
plot(banque$credcon)
```

## 2. Tableau disjonctif :

- (a) somme des lignes = nbre variables = 5

```
apply(tabdisj, 1, sum)
```

effectifs des colonnes ?

```
apply(tabdisj, 2, sum)  
colnames(tabdisj)
```

Modalités rares :

interdit : oui, crédit habitation : oui , crédit conso : fai fao, age : &lt; 25

## 3. ACM avec factoMineR :

- (a) les 3 graphes sont : 1) carte des catégories, 2) carte des individus 3) carte des corrélations au carré entre variables et composantes.

- (b) inertie et inertie relative de chaque axe

```
sum(resMCA$eig[, 1])  
resMCA$eig[, 1]
```

- (c) à partir de la stat du
- $\chi^2$
- :

```
chisq.test(tabdisj)$statistic / sum(tabdisj)
```

- (d) contributions principales des catégories sur l'axe 1 :

```
ctrs <- resMCA$var$contrib[, 1]  
ctrs[rev(order(ctrs))]
```

crédit habitation, femme, age &lt; 25

- (e) sur l'axe 2 :

```
ctrs <- resMCA$var$contrib[, 2]  
ctrs[rev(order(ctrs))]
```

interdit oui, age &lt;35, age &lt;75

- (f) axe 1 : clients avec crédit habitation, gros crédit conso, hommes, âges actifs (45,55) (= très consommateurs de services bancaires) contre jeunes, femmes, retraites, en interdit bancaire (peu consommateurs)

axe 2 : clients à risque (interdit, age &lt;35, petit crédit conso) contre peu risqués (pas d'interdit, retraités, gros crédit)

- (g) Utilisation de variables supplémentaires :

→ axe 1 : assurance vie, agriculteurs à droite(consommateurs), versus étudiants, inactifs, à gauche (peu consommateurs)

→ axe 2 : ouvriers, soldes négatifs (n1) en haut, compte d'épargne en bas.